Evolutionary Debunking, Realism and Anthropocentric Metasemantics[1]

Mark van Roojen

1. Introduction

Evolutionary debunking arguments can be unnerving. If successful, they show an incompatibility between two commitments that are pretty deep in my world view. I'm rather sure that I know that certain sorts of actions are wrong. And I'm sure that we are as we are as a result of certain natural evolutionary processes, at least roughly of the sort Darwin hypothesized. So it should be a comfort to agree with prominent debunkers such as Sharon Street who claim that such arguments favor anti-realist or constructivist theories over completely mind-independent accounts of morality. It already strikes me as implausible that the nature of morality would have nothing at all to do with the nature of human agents. If Street is right, my minimally constructivist sympathies would insulate my views from such debunking. I could then just heave a sigh of relief and be grateful that my own commitments are not in tension.

As it stands, however, I think these debunkers are wrong to see their arguments as raising a special problem for realism. In a nutshell, I think they conflate mind-dependent content-determination relations with mind-dependent content. Or to come at the point from a different direction, insofar as reference is broadly an epistemic relation, evolutionary debunking arguments would cause trouble for mind-independent theories of reference and content-*determination*, since there would be no guarantee that reference would track epistemic access. But a firmly realist theory of content is consistent with a mind-dependent theory of reference and

content-determination.[2] To use a toy example, most facts about rocks are mind-independent, but we are only in a position to refer to and talk about rocks because of the faculties that enable us to know about rocks. This should seem platitudinous, but it turns out to have important implications for evolutionary debunking arguments. These are (1) that fully realist externalist naturalist views are in no real danger from these arguments. (2) That a certain sort of non-naturalist view might also be able to use a parallel strategy to resist debunking, but (3) this will be somewhat harder to do given one of the thoughts that often motivates non-naturalism. I will argue for these claims by presenting a model for a realist theory of content that seems untouched by evolutionary debunking. Finally I express a worry that internalist views of the sort I myself favor may have problems similar to those I raise for non-naturalism, whether they are naturalist or not. But first some setup.

2. Debunking Deployed Against Realism

I'll begin with a sketch of a relatively generic[3] evolutionary debunking argument, presenting it in an order I find intuitive:

> 1. The range of moral judgements we actually make is significantly dependent on our affective responses and, in turn, on our capabilities and dispositions to have such responses. (Affect effects judgements.)

> 2. Our affective capabilities and dispositions are as they are as a result of a process of

---

[2] A point very well made in a more general way in Schroeter & Schroeter (ms).

[3] In this generic form it is in the ballpark with arguments by Street (2006, 2008, & 2012) and Bedke (2009, 2014), and perhaps Joyce (2006, 2016). My discussion below will focus more on Street's views since her targets are all response-independent views, whereas Bedke just aims at non-naturalist views. I'll say something about how such arguments might be resisted by non-naturalists at the end of the paper.

random variation, culling and inheritance. The culling process tended to allow dispositions to be inherited only if they were relatively advantageous for survival in the environment where they occurred. Relatively disadvantageous variations tended to go to extinction through the same process. (Affective capacities evolved.)

3. Therefore, if our moral judgements come anywhere close to tracking the truth, it must be because this process of variation, culling and inheritance shaped our affective dispositions so as to allow our judgements to do so. (Truth tracking must be selected for.)

4. If the moral truths are independent of our psychological capacities and dispositions (or of adaptiveness in our environment), it is highly unlikely that such a random process would have caused us to have capacities and dispositions that allow our judgements to come relatively close to tracking those truths. (No selection to track affect-independent truths.)

5. Therefore, we should conclude either that we are likely not coming relatively close to tracking truths with our moral judgements or that the moral truths we track are mind-dependent in some significant way.

The argument has three premises, steps 1, 2 and 4. The first two premises, properly understood, are true. The third step, properly understood, follows from the first two. Since I dissent from the conclusion I must either dissent from the remaining premise, step 4, or argue that the conclusion does not follow from it. I think there is a reading of the fourth premise on which it is true, but that the conclusion does not follow from it. There is also a reading from which the conclusion does follow but that reading of 4 is not in fact true. The former deploys an antecedent probabilistic notion of likelihood, whereas the latter reading is about current epistemic

probability. I'll leave that cryptic claim alone for now, coming back to explain it after I sketch my response on behalf of realism later in the paper.

I should probably explain how I understand the two premises I accept full stop and why they're true so understood. The first postulates a causal influence running from our affective dispositions to the range of moral and other normative judgements we actually make. It does not say this is the only influence, or that these dispositions by themselves determine us to make the judgements we do. But it does imply a certain counterfactual dependence—if we lacked some of the affective capacities and dispositions we have, we would not make a wide and important range of the judgements we now make and believe. For example, without some empathetic understanding of other people and other creatures we would be unlikely to endorse altruism to the extent we do. Without a disposition to value reciprocity, we would be unlikely to judge friendship an important value and to find fairness important. And so on. So understood, the first premise should be uncontroversial. The second premise recognizes that these capacities are to a large extent contingent and that we have them as the result of evolution which at one level involves random variation. To the extent that there is selection among traits, it is not intentional selection. Rather, it is a process where the traits which are more disadvantageous than alternative traits present in a species lead creatures with those traits not to do as well as those without them, and hence to leave fewer progeny with the disadvantageous trait. The traits that are more advantageous thus have a greater tendency to persist. The upshot is not that traits which persist are optimal, only that they were more adaptive than the available alternative thrown up by random variation.

Taken together these claims imply that any affective dispositions that enable us to track the truth about morality must be the result of this process of variation and filtering. That's the third step in the argument, which follows from the first two. Sharon Street, most prominently,

4

has defended the fourth step (and third premise) of the argument. She argues that realists about value (as opposed to anti-realists) cannot explain the congruence between "our evaluative attitudes, on the one hand, and the independent evaluative truths that realism posits, on the other" (2006: 109). Bedke (2009, 2014) similarly argues that any congruence would be a cosmic coincidence. We'll need to look at Street's more detailed argument, but first we need to get clearer about the realism/anti-realism divide. Everybody thinks that some moral truths are attitude-dependent. One reason one shouldn't insult people is that it makes them feel bad. Street suggests we can get a clean divide between realists and anti-realists if we look at whether a theory posits any judgements whose truth or falsity is determined independently of the whole set of evaluative judgements we do make or might make upon reflection as well as any other evaluative attitudes we hold or would hold upon reflection. Realists will think that there are moral truths whose truths do not depend on these attitudes, whereas anti-realists in her sense will think all moral truths have this sort of dependence (2006: 111). This way of dividing realism from anti-realism is supposed to come to the same thing as dividing "stance-independent" realist truths from "stance-dependent" irrealist claims (Street 2006: 111; Shafer-Landau 2003: 15; Milo, 1995, 182). So understood, it looks like certain sorts of naturalists—those who don't reduce moral truths to facts about human psychology—and non-naturalists will count as realists.

Irrealists, Street suggests, will have an easy time explaining the congruence of moral truth with the judgements that our affective capacities and dispositions allow us to make. Rather than explaining why our capacities evolved so as to allow us to track an independent set of truths, they can reverse the order of explanation. They can argue that insofar as our attitudes enter constitutively into the nature of the truths in question, it is no surprise that those truths will co-vary with the relevant attitudes. So, for example, if the moral truths just are the judgements we would stably make under conditions of full information and imagination including empathy,

5

it will be no surprise that when we have such information and are sufficiently imaginative and empathetic, we'll have some idea of what is in fact right and valuable.[4]

Clearly realists won't be able to give this kind of explanation. There will be no non-trivial condition under which our judgements about value must be right. So they cannot define rightness in terms of those judgements to secure the requisite correlations. But that is not to say that they might not have other explanations available. In fact, as I will soon argue, there is a perfectly good explanation of how evolution could put us in a position to know about realism-friendly mind-independent normative properties. If evolution in fact does that, the reading of premise 4 that supports the conclusion in my reconstruction of the antirealist argument will be false.

But it isn't obviously false, so we should look at some arguments to the effect that premise 4—the claim that realists cannot explain the congruence of our inherited evaluative dispositions with moral truth—must be true. In her original presentation of her argument, Street suggests that either natural selection of evaluative dispositions tracked the moral truth, or it didn't. If it didn't, then while occasional moral judgements might by chance turn out to be true, there would be no general tendency for our moral beliefs to be close to true. Even when corrected by deliberation and reflection, evaluative dispositions which are only ever aligned with the moral truth by chance will be no likelier to help us get things right as to get things wrong. That's because these processes will have at least as many misleading as veridical judgements to work from (2006: 124).

On the other hand, if there is some relation between the affective tendencies favored by evolution and the independent normative truths, and if that relationship is positive rather than

_____

[4]

 Tropman (2014) makes the case that the advantage to irrealism/constructivism is illusory by arguing that such views have their own problems.

negative, there might be some hope of vindicating the judgements made available to us by evolution. Street suggests that the required relation would have to be one in which we evolved the abilities to make true judgements because true judgements contribute to reproductive success. She thinks that the true judgements would have to be conducive to reproductive success "because they are *true*, and it proved advantageous to grasp evaluative truths" (2006: 128). The problem for realists on this horn of the dilemma is that there is a better explanation of how we came to make these judgements that doesn't posit their truth. This explanation suggests that creatures who believed (some of) the judgements that our evaluative dispositions enable us to make tended to act in accordance with those judgements. And acting in accordance with those judgements tended to favor reproductive success (2006: 129). Absent some showing that this tendency is itself to be explained by the moral truths, we have a more parsimonious explanation that does just as well explaining these dispositions.

There is at least one widely noted problem with this argument—it isn't obvious that realists need to say that the ability to make true moral judgements was adaptive because those judgements are true. There could be a non-accidental connection between the ability to make judgements that are as a matter of fact true and reproductive success in an environment. This could favor creatures who are able to make such true judgements, but not because they are true. So-called "pre-established harmony" or "third factor" (Enoch 2010) accounts posit just such an explanation. Survival is good as is reproductive success, and not accidentally so (Enoch 2010: 168ff.). Creatures with certain cognitive capacities necessary to being able to make moral judgements necessarily have rights (Wielenberg 2010). Well-being is good (Brosnan 2010). Pleasure is intrinsically good (Skarsaune 2011). And it is not an accident that evolution would select for the belief that survival and reproductive success is good, or that creatures like us are protected by a moral barrier, or that pleasure is good, or… Defenders against evolutionary

debunking can claim that it is not at all implausible that we evolved to have these very beliefs, so that given the necessary truth of these moral claims it is not an accident that we believe the truth. I think there is something to these replies,[5] but there remain some worries. One has to do with parts of morality that don't seem explicable in this way. For instance, it isn't obvious that being right about survival, reproductive success and rights is enough to generate any particular view about our duties to animals. Another worry has to do with the thought that knowledge requires believing something because it is true,[6] and these explanations don't seem to secure that. Still, I'm not dismissing this strategy of response, partly because I'm not sure my own actual views can take advantage of the response strategy I am about to offer the thoroughgoing realist.

3. A Strategy of Realist Response

The thought that realism cannot explain how we come to accurately represent moral reality given the stance-independence of moral properties seems very compelling until one distinguishes the nature of the properties themselves from the semantic theory which explains how we come to be able to talk about these stance-independent properties. The properties we refer to can be perfectly objective stance-independent properties while our ability to refer to these properties can depend on all sorts of contingencies about our natures. I will argue for this claim by giving a model from another domain that has both of the features I claim a realist theory of morality could have. With that example in place, I will show how it is compatible with thinking it antecedently unlikely that we came to have knowledge of these very properties, while still thinking that it is now very plausible to think that we have such access.

---

[5] I used selection for limited altruism to make a similar point in a 2006 APA symposium on what became Street (2011) .

[6] Whatever that elusive "because" comes to. Discussions of sensitivity and safety (for example Clarke-Doane 2011, 2012) strike me as attempts to partly elucidate the notion, but I suspect there is more to it.

My model comes from a theory of colors developed by David Hilbert (1987). He calls the view "Anthropocentric Realism."[7] On this view, colors are objective observer-independent properties of objects, roughly their surface spectral reflectances. Color "is the disposition objects have to reflect varying percentages of incident light" (1987: 119). That an object has such a disposition is completely independent of observers, whether actual or possible. It likely depends on the microstructure of the surface and the nature of light, but not on anything subjective. Therein lies the view's realism. The anthropocentrism comes in to explain which such dispositions we refer to with our color terms. A first level of anthropocentism is present insofar as different languages divide color space differently from one another (1987: 130). We can discriminate many more colors than we have words for. Our color terms pick out classes of similar reflectance properties that look similar to us, but where one class begins and another one ends is somewhat a matter of choice. Once the boundaries are decided, however, the classes themselves are perfectly objective. As Hilbert notes, we could design a device to sort token color samples into the classes. The groupings are to some extent up to us, but that doesn't mean that our terms don't pick out real properties that would be instantiated or not whether we had existed or not.

A second level of anthropocentrism comes into the picture because we have three kinds of color receptors, each sensitive to light within a certain range. Colors are computed on the basis of the relative intensities of light in the ranges to which these receptors are sensitive. This means that two surfaces with different reflectance profiles will look the same to us across a range of normal lighting circumstances, so long as the peak reflectances are indiscriminable by each sensor. This is the phenomenon of metamerism. "Metamers do not differ just slightly in

[7] This is far from the first metaethics paper to exploit an analogy with color. Phillip Pettit's 2015 Tanner Lecture (forthcoming, 215), for example, deploys it to point out something related to my point - that anthropocentric properties can be both natural and real.

their reflectance profiles. An object with an essentially flat reflectance curve can appear to have the same color as one with large peaks in reflectance" (1987: 103). Our color terms, then, will group reflectance patterns together as cardinal red (say) just when they affect our receptors in the same way. And in one sense very different reflectance properties can do that. Still, Hilbert suggests, we can offer a unifying account of the relevant properties of a sort:

> … human color vision is … a passive system with three types of broad-band sensors. As we have seen, objects that reflect the same amounts of light in each of the wave-bands will appear to have the same color. We can express the common property that the reflectances of such objects will share by summing the reflectances of the objects over each of the three ranges. We will obtain what is called a triple of integrated reflectances. (1987: 111)

Once again these are real properties that would be there whether we existed or not. But our talking about them is a function of what we are like. Why does our color talk refer to triples of reflectances at just these bands? Because our visual system has sensors that are sensitive at just these. Other creatures could and probably do have visual systems like ours but sensitive at different bands. There might be creatures with more or fewer broad-band receptors. If there were such creatures and they could talk about their visual experiences, they would according to this theory refer to different reflectance properties than we do. But for all that the reflectance properties that would be the referent of their "color" terms would be perfectly objective. We are presently surrounded by them. We just don't discriminate them as those creatures would and hence are not (usually) in a position to talk about them.

My use of the model does not really depend on its correctness as an account of our color terms, though I find it pretty plausible. It presents a model for stance-independent realism that makes it less mysterious how we could be in a position to get the facts right without introducing

any stance-dependence into the nature of the properties picked out. Hilbert's color realism identifies those properties with the perfectly objective physical properties to which we are sensitive. But our being sensitive to them is not part of the nature of the properties themselves, nor are those properties in any way dependent on us. However, the story of how we come to be in a position to talk about these properties—how our color terms come to pick out just these properties—does depend on our nature. This means that the semantic values of the color terms just are mind-independent properties. But the metasemantic explanation of how our terms come to have these referents makes essential reference to our ability to discriminate these properties and that means that the full story will get into the contingent details of our visual system. On a Millian version of the metasemantic story, we would begin with some population of people who are able to distinguish the relevant reflectance properties by how they look to them and who coin terms to pick out certain classes of them. These are then passed on to a wider community of users, who use them to refer to the same properties as the original bunch did. On this story, the actual ability to talk about these properties depends on the nature of the observers, but the properties themselves don't. And, since it is the properties that are picked out by the color terms, there is no anthropomorphism in the contents of sentences using color vocabulary. A yellow rose would be and remain yellow, even if our visual capacities and dispositions changed. We would rather lose our ability to discriminate yellow from other colors.

There is a lot of evolutionary contingency in the story. Probably the exact bands of light to which our receptors are sensitive could easily have been otherwise, and we might have had more or fewer receptors (though not many more or many fewer).[8] When our sightless ancestors started the process of evolving into us, it was a long shot that we would see the colors we do. It

---

[8] Most non-primate mammals have two; reptiles, amphibians and most birds have four; and pigeons have five. https://en.wikipedia.org/wiki/Color_vision#In_other_animal_species.

is improbable that the exact shades we do in fact see are such that we were selected to see these and not some other set within the range that is illuminated on earth. So there is no tracking explanation for our seeing just these colors rather than some other colors.[9] Yet that should not make us skeptical of our ability to see the colors we see. It is no miracle that we are sensitive to surface spectral reflectance, and being sensitive to it will require some color perception system or other. Any one of these would have been a long shot. So it should be no surprise that the one we wound up with was one of these long shots.

4. A Similar Model for Value Realism

Realists about values and morality of the sort Street targets think our moral terms function to pick out perfectly objective stance-independent properties. I take it that most of the Cornell realists qualify. I will use Richard Boyd's (1988) account as a basis to sketch a view of value parallel to Hilbert's views about color. Boyd thinks that moral goodness is a homeostatic cluster property—that is a property composed of a number of properties which, when co-instantiated, tend to keep each other in existence. The model for these properties is health in animals including human beings. Health has a number of components and a healthy creature is such that the presence of one component tends to reinforce the presence of the others. These properties are perfectly stance-independent. They would be instantiated where they are, whether or not we had ever come to think and talk about them.[10] Reference to a value property is secured by an epistemically virtuous feedback loop, leading from the property to what we think and say about it and to how we interact with it, which in turn provides us information we can use to modify our

---

[9] There's not really an "adaptive link" explanation for that either, if I'm right that some similar but different set of receptor would have served us just as well where survival is concerned.

[10] Boyd suggests that with all scientific properties it is a matter of luck whether we come to be in the appropriate relation to refer to them.

views of the property. Boyd emphasizes the epistemic nature of the relation:

> *Roughly*, and for nondegenerate cases, a term *t* refers to a kind (property, relation, etc.) *k* just in case there exist causal mechanisms whose tendency is to bring it about, over time, that what is predicated of the term *t* will be approximately true of *k* (excuse the blurring of the use-mention distinction). Such mechanisms will typically include the existence of procedures which are approximately accurate for recognizing members or instances of *k* (at least for easy cases) and which relevantly govern the use of *t*, the social transmission of certain relevantly approximately true beliefs regarding *k*, formulated as claims about *t* (again excuse the slight to the use-mention distinction), a pattern of deference to experts on *k* with respect to the use of *t*, etc… When relations of this sort obtain, we may think of the properties of *k* as regulating the use of *t* (via such causal relations), and we may think of what is said using *t* providing us with socially coordinated *epistemic access* to *k*: *t* refers to *k* (in nondegenerate cases) just in case the socially coordinated use of *t* provides significant epistemic access to *k*, and not to other kinds (properties, etc.). (Boyd 1988: 195)

Whether this is a good account of reference in general or moral reference in particular is not presently my concern. What I want to note is that on this account we must have epistemic access to the properties that are the referents of our terms.

This means that according to this kind of realist theory the referent of our terms must be something we have epistemic access to. And that means that the reference-determination story will bring in a kind of stance-dependence, not of the property which is the referent, but of our ability to refer to it. I think this is a general feature of theories of reference, but for now I'll stick with my example. Suppose now that the realist takes the first horn of Street's dilemma, and suggests that evolutionary processes did not select for correct moral judgements. She suggests:

… [W]e are left with the implausible skeptical conclusion that our evaluative judgements are in all likelihood mostly off track, for our system of evaluative judgements is revealed to be utterly saturated and contaminated with illegitimate influence. We should have been evolving towards affirming the independent evaluative truths posited by the realist, but instead it turns out that we have been evolving towards affirming whatever evaluative content tends to promote reproductive success. We have thus been guided by the wrong sort of influence from the outset of our evaluative history, and so, more likely than not, most of our evaluative judgements have nothing to do with the truth. Of course it is possible that as a matter of sheer chance, some large portion of our evaluative judgements ended up true, … but this would require a fluke of luck … (2006: 122)

This is not what we should say if we accept Boyd's reference-determination story. Were we in a scenario where most of our beliefs about the relevant evaluative properties were false and if we lacked a feedback mechanism that allowed us to correct these false beliefs over time, we would not be referring to these properties with our evaluative terms. We would either be referring to a different set of properties—properties which do stand in the requisite epistemic feedback relation to our use of those terms—or we would be referring to no properties at all.

Hilbert's color theory provides a parallel. It is unlikely that we were selected for our ability to see fuchsia. It is also unlikely that at the beginning of sighted existence our current visual apparatus was anything more than one of many ways that creatures might implement a light sensitive perceptual system. More probably, any one of a range of possible implementations was a real possibility[11] and the accidents of our evolutionary history explain why we have the one we do. But we have the one we have and we can discriminate fuchsia. Suppose things went

---

[11] What range? My not very educated guess is that the most probably systems would have had more than two but not more than five broad band sensors and that the sensors would have all been sensitive to some range of the light actually present in our environment. But the specific boundaries of these sensors would be somewhat up for grabs.

14

differently, so that our actual visual system did not group all and only the reflectance profiles we see as fuchsia together. This would be a world in which some things were still fuchsia since there would still be the reflectance profiles picked out by our present term 'fuchsia'. But we would not see colors in this profile as similar to one another. And we would not have a word for fuchsia and would not talk or think about fuchsia. If we used a term phonetically and orthographically like 'fuchsia', it would pick out something other than fuchsia.

This does not mean that we now as we are have benefited from some tremendous fluke of luck to be able to get things right about fuchsia, about which we would otherwise have been wrong. True, the antecedent chances of our developing just this visual system was very small. The discriminatory abilities which underwrite the epistemology of color also secure our ability to refer to the colors we do. The contingent fact that we evolved to have these abilities is responsible for both our ability to talk about fuchsia, and our knowing some things about it. That we can do both rather than only one is not a fluke. What is in some sense a fluke is that we perceive precisely fuchsia, and the other colors we see. But whatever colors we would have been in a position to discriminate, it was antecedently a long shot that we would evolve to be able to know about them.

Boyd should say the parallel thing when it comes to the dispositions that underwrite our evaluative judgements. Yes, it was not necessary or perhaps even likely that we wound up in a position to grasp the evaluative truths we recognize. We might have been different. But then we would not have been talking about the same evaluative properties we presently do talk about.[12]

---

[12] Shafer-Landau (2012: 11) makes a suggestion that is similar but different: "[C]ertain moral propositions are reference-fixing, such that denial of (enough of) these propositions shows that an agent is no longer talking about morality at all." His thought is that people who make different judgements about certain moral paradigms won't exercise "semantic competence" and hence won't be talking about morality. This makes the getting these things right a "conceptual truth" about the senses of the relevant moral terms. It is compatible with Boyd-style metasemantics that people believe all sorts of crazy things about the referents of their terms, just so they have an epistemic pathway to learning the truth. FitzPatrick

We might have been talking about some other properties, perhaps similar enough to our present properties that we should think of them as evaluative as well.

At this point it would be natural to respond that this is a form of relativism and that relativism is a form of anti-realism, not realism. Street in fact says something like this:

> In order to count as genuinely realist, then, a version of value naturalism must take the view that *which* natural facts evaluative facts are identical with is independent of our evaluative attitudes. For ease of expression, let us put the point this way: in order to count as realist, a version of value naturalism must take the view that facts about *natural-normative identities* … are independent of our evaluative attitudes. (2006: 137)

I think that this is a mistake, though an easy one to make. The postulated theory does not say that the truths would change depending on which attitudinal dispositions we had evolved to have. It only says that our present sentences would not have picked out the same propositions under alternative scenarios. Most of the same propositions would remain true (though not the ones about our evolutionary history). We would just be ignorant of them. The point is perfectly general. If we never had epistemic access to the things we think and talk about, we would not have been able to talk about them. This holds for evaluative things, color, the Big Bang, rocks and trees. This isn't relativism, it is just an ordinary fact about reference-determination that falls out of Boyd's metasemantic theory.[13]

Relatedly, when this kind of naturalist identifies evaluative properties with natural properties and thus evaluative facts with natural facts, these identities are not supposed to be creatures of our thought or language. Everything is necessarily self-identical. It is possible that in

___

(2014) and Vavova (2014) make similar suggestions to Shafer-Landau's.

[13] See Schroeter & Schroeter (ms: n. 21) for the same point pressed against Street.

some sense we might have picked out different properties with our present evaluative terms and perhaps these properties would have been similar enough to those we presently use our terms for that we should call those evaluative as well. Nonetheless, we would be talking about different properties and facts than we presently do—at least if something like this theory of reference determination is correct.[14]

At this point it might be very tempting to pull out Horgan and Timmons's (1992) Moral Twin-Earth (MTE) argument to argue that we would take our counterfactual evolutionary counterparts with different evaluative dispositions to disagree with us when they make judgements consistent with their less empathetic moral dispositions.[15] The thought experiment is after all aimed at Boyd's semantic and metasemantic theory since it predicts that there will be no real disagreement when our judgements are appropriately causally and epistemically related to different referents. There are several things to say here. The first is that semantic externalists in broadly the same camp as Boyd wish to deny the probative value of those thought experiments (Schroeter & Schroeter 2014, Dowell 2016,). Secondly, those thought experiments are themselves viewed by their proponents as bringing in the kind of stance-dependence that Street thinks her argument from evolution shows we need. Horgan and Timmons use the argument precisely to argue for a species of non-cognitivism according to which all moral judgements must be made from an affectively engaged stance. If they're right, it isn't obvious that the evolutionary argument is now adding anything.

Furthermore, it is a delicate matter exactly what the Moral Twin-Earth argument shows.

14

 I think similar results would fall out of most theories of reference-determination, but I won't argue for that here.

15

 This is in fact what Street does when she considers the possibility of "rigidified" naturalism.

I think it does support a version of moderate morals/motives internalism.[16] But it does not, I think, show that semantic externalism of roughly the sort Boyd proposes is incorrect.[17] If I'm right, realists who think the MTE argument is no reason to abandon their semantic theories, have no other reason not to avail themselves of this response exploiting the epistemic character of reference-determination. And if they do so avail themselves, they will have a metasemantic theory which makes it the case that moral terms refer only to those things to which they stand in an appropriate epistemic relation. On the other hand, if the MTE scenarios really show that no externalist metasemantics is possible, and if only such metasemantic views deliver the upshot that reference is a kind of epistemic relation, my overall point would be interesting but ultimately of no use to a realist.

5. Connecting the Response to the Original Debunking Argument

I said earlier that the fourth step of the generic debunking argument with which I began had a true reading that did not support the conclusion and a false reading which did. I can now explain. That premise was:

> 4. If the moral truths are independent of our psychological capacities and dispositions (or of adaptiveness in our environment), it is highly unlikely that such a random process would have caused us to have capacities and dispositions that allow our judgements to

---

[16] Moderate morals/motives internalism holds that it is necessary that in some range of normal cases moral judgements motivate those who make them. See Dreier (1991) and van Roojen (2010, 2017) for the connection with the MTE.

[17] Boyd's theory can be modified to take the MTE argument into account while preserving the referential nature of the semantics. See van Roojen (2006). It would be too difficult to work what I say there into this paper, but it may be important to note that the problem for Boyd in the MTE story does not turn on a lack of epistemic access to the possible referents of their respective moral terms. Each community is in fact in causal and even epistemic contact with instances of both properties.

come relatively close to tracking those truths. (No selection to track affect-independent truths.)

It is true that we might easily not have been in a position to track truths involving these particular properties. That claim is true. But from this the conclusion stated in step 5 does not follow:

5. Therefore, we should conclude either that we are likely not coming relatively close to tracking truths with our moral judgements or that the moral truths we track are mind-dependent in some significant way.

It does not follow because, had we not tracked these particular truths, we would have been talking about something else about which we might still be in a position to know quite a lot. Given the way we evolved, the phenomena we are talking about are the phenomena our evolutionary history put us in a position also to know about; we can refer only to those things we have in our epistemic grip. So we are at least on track to know some things about values and morals. But the fact that *reference-determination* is mind-dependent does not show that the properties picked out in this mind-dependent way are mind-dependent.[18] So both options offered by five can be false, consistent with the true reading of 4.

6. Can Non-naturalist Realists Make Use of the Reference-Determination Strategy?

I've offered a response to evolutionary debunking arguments on behalf of naturalist realists. Can

---

[18] Again, a point well made in Schroeter and Schroeter (ms.). I should add that the point is pretty general since many theories of reference require that we stand in the right epistemic relation to the things to which we refer. As David Copp helpfully noted to me in response to this paper, "Could any term come to refer to a property in a way that did not depend on our ability to discriminate the property?"

non-naturalist realists adopt a similar response? I find it really hard to say, partly because the commitments of non-naturalism are pretty slippery and because it is hard to say what all and only the natural properties have in common and thus what commitments separate naturalists from nonnaturalists. Given the existence of non-reductive naturalists such as Sturgeon (2006), we might wonder what naturalists and non-naturalists are arguing about. Non-reductive naturalists and non-naturalists agree that moral properties are further properties over and above the natural properties on which they supervene. And they can agree that the relations of the higher-level normative properties of actions or objects to the lower-level properties of those actions or objects are very much like the relations of biological or other special science properties of objects and the properties of lower-level sciences like physics. Yet, so-called "robust" realists like Enoch (2011) seem to differ from naturalists mostly by insisting that the relevant properties are "just too different" from ordinary natural properties to be of the same general type. We might then see their disagreement as being over whether normative properties belong in a natural kind of natural kinds—the natural. If the natural properties themselves form a natural kind, there is some real similarity that they all share. Naturalists think that normative properties too share this real similarity and hence belong in this higher-level natural kind of kinds. Non-naturalists disagree because they think that normative properties are "just too different."

What then might these differences be? Some non-naturalists insist that such properties are simple, whereas others (Scanlon 2014, McDowell 1985) suggest that normative properties have a kind of relational structure. But simplicity doesn't seem to be the right kind of difference. Some perfectly natural properties are simple, and in any case non-naturalist adherence to that commitment seems to be a historical accident stemming from Moore's (1903) use of simplicity to explain unanalyzability along with his conflation of the nature of our thought about the

20

properties with the nature of the properties themselves. Many non-naturalists deny the causal efficacy of non-natural properties, and Bedke takes this as constitutive of non-naturalism in his debunking argument.

Denying causality to non-natural properties would seem to rule out a flat-footed application of Boyd's causal regulation semantics. But it isn't obvious to me that non-naturalists should abandon direct reference accounts that depend on some sort of epistemic feedback from the property. First off, if the worry is that normative properties are Platonic universals and that such universals can't cause anything (Enoch 2010), Platonism is a very general view and if it is plausible, it would apply to all properties, natural and non-natural alike.[19] This version of the objection would show that no properties can cause anything, not something any view should take on board. In any case, all Boyd's theory really needs is that events in which the relevant properties are instantiated cause other events so as to give us evidence about the nature of the properties instantiated. This may suggest a different objection: perhaps the thought is that normative properties are epiphenomenal, supervening on underlying properties which do the real causal work. But the natural properties favored identified with moral properties by naturalist moral realists also supervene in just this way. Homeostatic cluster properties supervene on and are constituted by the properties physics typically works with. The causal impotence of higher-level phenomena would undermine both naturalist and non-naturalist causation and hence causally mediated reference-determination. The views seem to be on a par. It would be more plausible in any case to follow common sense and let higher-level phenomena inherit the causal powers of the things that compose them. Baseballs can break windows because their molecules can. Mental states can cause action because the physical states that realize them can, and so on.

---

[19] I suppose one could be a Platonist about ethics but not about other properties, but I can't see why one would hold that view.

But if non-naturalists take that view, they should give up on causal impotence, as Russ Shafer-Landau (2012) seems willing to countenance.

So it would seem non-naturalists too could adopt a strategy parallel to the one offered to naturalist realists. They too should say that had we evolved to make sufficiently different moral judgements than we actually do, we would have epistemic access to different properties and hence be talking about those different properties. This requires the possibility of different non-natural properties as possible referents for the judgements made in those counterfactual circumstances. But without saying more about what makes these properties non-natural, I don't see any in-principle reason there could not be. If different higher-level natural properties can supervene on lower-level physical properties and if therefore instances in which they are instantiated give us causal-epistemic feedback, we would have a picture parallel to the naturalists' view, substituting non-natural for supernatural properties.

But this doesn't seem to me to be much in the spirit of many if not most non-naturalist views. I think there's another motivation for the "just too different" thought. And this motivation would not sit comfortably with allowing that had we evolved differently we would have referred to different properties with our moral and evaluative terms. My hypothesis is that many non-naturalist think that the connection such properties have with reasons—their normativity—makes moral and evaluative properties special, and that none of the natural properties have that same connection. One way to cash out the thought is as a kind of morals/reasons existence internalism[20]—moral and evaluative properties would have to be necessarily reason-giving at

---

[20] Such internalism falls immediately out of a view like Scanlon's which just works with reasons as the basic normative notion. It is not as clear whether other prominent non-naturalists like Enoch or Shafer-Landau would endorse it. But still it seems to me as one of the better ways to try to cash out the alleged differentness of the normative, though other morals/reasons internalists are happy to embrace naturalism.

least to us. If we couple that idea with the suggestion that a sufficiently different evolutionary endowment would have caused our evaluative and moral terms to track different properties than they actually do, we are faced with an obvious question. Would those different properties have been genuinely normative? In other words, would they be reason-giving? If the answer is no, we do, I think, wind up back in the grip of a worry about coincidence. How are we so lucky as to be the ones who track the only genuinely normative properties around? (And similarly, how are we so lucky as to be tracking the only non-natural properties around?) If we take the other option and say that those properties too are genuinely normative and reason-given, we once more see an obvious question. Are they reason-giving to one and all, or just to people with these alternative evaluative dispositions? If the answer is "to one and all," we will likely have conflicting reasons stemming from the two sets of normative properties with non-tracking extension. That doesn't seem like an attractive view. Better to say that each property is normative for those with the disposition that allowed their judgements to track those properties. This looks a lot like stance-dependent normativity. And now the non-naturalist view winds up not being realist in Street's sense.[21] And yet it looks like the cosmic coincidence worry has some force for those who hold it.

Cornell realists don't have the same issue because they are already quite happy to give up on morals/reasons internalism. Their strategy has typically been to deny that moral properties are necessarily reason-giving (Brink 1989). They have instead attributed a weaker status to moral properties, that of being interesting to us much of the time. Their strategy has generally been to explain why internalism was plausible without postulating a necessary connection between

---

[21] It is in fact striking that Street (2008: 214ff.) brings in a dilemma involving morals/reasons internalism to answer Copp's (2008) stance-independent realist response to her challenge. She suggests that only a reasons internalist reading of this theory is genuinely normative, but then argues that this kind of theory has trouble explaining congruence. If the theory is reasons externalist it isn't normative, she suggests. That's in essence the same dialectic as is going on here for the non-naturalist. If your theory captures the motivations that make internalism attractive, you have trouble retaining a stance-independent view.

reasons and moral properties. If humans would often enough be motivated by goals that were attractive because of their moral attributes, externalists could agree that these properties are reliably connected with our reasons and motives without taking that connection to be necessary (Railton 1993). Were we to have evolved differently, so that our evaluative dispositions caused us to fix on other properties as most relevant for practical purposes, we would as the debunkers argue, have made different moral judgements. But we would also have valued different things so that those things would in that case also have the reliable but not necessary connection with action that the morally valuable things now actually do. Or so the Cornell realists should argue.

7. Where Are We and How Did We Get Here?

We have seen that a property can be completely objective and stance-independent in the sense that it would exist whether we were in a position to know about it or not even while our ability to think and talk about it does require us to be in a position to know about it. David Hilbert's theory of color presents a model for such properties and reference to them. And we have explored how an externalist metasemantics for moral terms like Boyd's causal regulation theory can exploit this possibility to explain how we in fact manage to refer to such properties despite it being unlikely, antecedent to our actually having evolved to refer to these particular properties, that we would wind up referring to precisely those properties we do. Furthermore, since our being able to think and talk about them depends on our being in the relevant favorable epistemic position with respect to these properties, such a view eludes the argument from the unlikelihood of our having come to have just the dispositions that put us in this favorable position to the

unlikelihood of our now getting things right. Under the relevant skeptical scenarios we would have been talking about different properties about which we would then be getting things right. So Street is wrong to think that stance-independent realist views are especially subject to her debunking argument.

Nothing in the Boyd metasemantics places all that many conditions on the nature of the properties referred to, beyond their being able to generate epistemic feedback through some causal powers or other. This opened up the possibility that non-naturalists willing to countenance causally efficacious non-natural properties could adopt a similar metasemantics and answer evolutionary debunking arguments in parallel ways. In fact, I claim, this possibility is real. But it is in tension with one motivation for non-naturalism. This motivation is that normative properties are special - too special as the nonnaturalists think of it to be natural. There are different ways to cash this out, but one way might be to say that they are the only properties that could be genuinely reason-giving.[22] But if we say that, then had we tracked any other properties with our normative terms they would not have been genuinely normative. Only some epistemic feedback loops would put us in touch with genuinely normative properties. And now this non-naturalist is again saddled with the cosmic coincidence worry that the Cornell Realist was able to parry, for we have no explanation of how come we came to be so lucky as to track the only genuinely normative properties.

That's where the argument has taken us so far. But we should notice that non-

---

[22] Or more weakly, any property that had (for example) racism or some other intrinsically wrong practice in its extension could not give us reason to go in for such things, in the way that the rightness of anti-racism gives us reason to be against racism.

naturalness itself has played no real role in the argument. What did the work was the idea that the properties we track are genuinely normative and reason giving, whereas the properties we would have tracked had our randomly selected propensities been different would not have been genuinely normative. A nonreductive naturalist might be tempted by that thought as easily as a non-naturalist. As might someone like myself, who is currently agnostic between these nonreductive options. I admit to being tempted, but this line of thought has me thinking twice.

References

Bedke, M. 2009. "Intuitive Non-Naturalism Meets Cosmic Coincidence," *Pacific Philosophical Quarterly* 90: 188–209.

———. 2014. "No Coincidence?" In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics* 9. Oxford: Oxford University Press: 102-125.

Boyd, R. 1988. "How to Be a Moral Realist." In G. Sayre-McCord (ed.), *Essays on Moral Realism*. Ithaca: Cornell University Press: 181-228.

Brink, D. 1989. "Externalist Moral Realism," *Southern Journal of Philosophy*, 24, Supplement: 23–40.

Brosnan, K. 2011. "Do the Evolutionary Origins of Our Moral Beliefs Undermine Moral Knowledge?" *Biology and Philosophy* 26: 51–64.

Clarke-Doane, J. 2012. "Morality and Mathematics: The Evolutionary Challenge," *Ethics* 122: 313–340.

———. 2014. "Moral Epistemology: The Mathematics Analogy," *Noûs* 48: 238–255.

Copp, D. 2008. "Darwinian Skepticism about Moral Realism," *Philosophical Issues* 18: 186–206.

Dowell, J. L. 2016. "The Metaethical Insignificance of Moral Twin Earth." In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics* 11. Oxford: Oxford University Press: 1-27.

Enoch, D. 2010. "The Epistemological Challenge to Metanormative Realism: How Best to Understand It, and How to Cope with It," *Philosophical Studies* 148: 413–438.

———. 2011. *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press.

FitzPatrick, W. 2014. "Debunking Evolutionary Debunking of Ethical Realism." *Philosophical Studies* 172.

Hilbert, D. 1987. *Color and Color Perception*. Palo Alto: CLSI.

Horgan, T. and M. Timmons. 1992. "Troubles for New Wave Moral Semantics: The 'Open Question Argument' Revived," *Philosophical Papers* 21: 153–175.

Joyce, R. 2006. *The Evolution of Morality*. Cambridge: MIT Press.

———. 2016. "Confessions of a Modest Debunker." In U. Leibowitz & N. Sinclair (eds.) *Explanation in Ethics and Mathematics*. Oxford: Oxford University Press.

Kahane, G. 2010. "Evolutionary Debunking Arguments," *Noûs* 45: 103–125.

McDowell, J. 1985. "Values and Secondary Qualities." In T. Honderich (ed.), *Morality and Objectivity*, 110–129. London: Routledge.

Milo, R. 1995. "Contractarian Constructivism," *Journal of Philosophy*, 92: 181–204.

Pettit, P. forthcoming. *The Birth of Ethics*, Tanner Lectures delivered at the University of California Berkeley, April 2015.

Railton, P. 1993. "What the Non-Cognitivist Helps Us to See the Naturalist Must Help Us to Explain." In J. Haldane & C. Wright (eds.), *Reality, Representation, and Projection*.

Oxford: Oxford University Press: 279-300.

Scanlon, T. M. 2014. *Being Realistic about Reasons*. Oxford: Oxford University Press.

Schafer, K. 2010. "Evolution and Normative Scepticism," *Australasian Journal of Philosophy* 88: 471–488.

Schroeter, L., and F. Schroeter. 2014. "Normative Concepts: A Connectedness Model," *Philosophers' Imprint* 14 (25): 1–26.

———. (ms). "Can Metasemantics Resolve the Epistemic Challenge to Moral Realism?," https://www.academia.edu/2923803/Can_Metasemantics_Resolve_the_Epistemic_Challenge_to_Moral_Realism. (Accessed 1/1/2017).

Shafer-Landau, R. 2003. *Moral Realism*. Oxford: Oxford University Press.

———. 2012. "Evolutionary Debunking, Moral Realism and Moral Knowledge," *Journal of Ethics and Social Philosophy* 71: 1–37.

Skarsaune, K. O. 2011. "Darwin and Moral Realism: Survival of the Iffiest," *Philosophical Studies* 152: 229–243.

Street, S. 2006. "A Darwinian Dilemma for Realist Theories of Value," *Philosophical Studies* 127: 109–166.

———. 2008. "Reply to Copp: Naturalism, Normativity, and the Varieties of Realism Worth Worrying About," *Philosophical Issues* 18: 207–228.

_____. 2011. "Mind-Independence Without the Mystery: Why Quasi-Realists Can't Have It Both Ways" In R. Shafer-Landau (ed) *Oxford Studies in Metaethics* 6:1-32.

Sturgeon, N. L. 2006. "Ethical Naturalism." In D. Copp (ed.), *The Oxford Handbook of Ethical Theory* Oxford: Oxford University Press: 91-121.

Tropman, E. 2014. "Evolutionary Debunking Arguments: Moral Realism, Constructivism, and Explaining Moral Knowledge," *Philosophical Explorations* 17: 126–140.

van Roojen, M. 2010. "Moral Rationalism and Rational Amoralism," *Ethics* 120: 495–125.

———. 2017. "Rationalist Metaphysics, Semantics and Metasemantics." In K. Jones & F. Schroeter (eds.), *The Many Moral Rationalisms*. Oxford: Oxford University Press.

Vavova, K. 2014. "Debunking Evolutionary Debunking." In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics* 9. Oxford: Oxford University Press:-76-101.

Wielenberg, E. J. 2010. "On the Evolutionary Debunking of Morality," *Ethics* 120: 441–464.